

Modeling human-like intelligent image processing: An information processing perspective and approach

Emanuel Diamant

VIDIA-mant, P.O. Box 933, 55100 Kiriat Ono, Israel

Received 29 May 2007; accepted 30 May 2007

Abstract

We live in the Information Age, and information has become a critically important component of our life. The success of the Internet made huge amounts of it easily available and accessible to everyone. To keep the flow of this information manageable, means for its faultless circulation and effective handling have become urgently required. Considerable research efforts are dedicated today to address this necessity, but they are seriously hampered by the lack of a common agreement about “What is information?” In particular, what is “visual information”—human’s primary input from the surrounding world. The problem is further aggravated by a long-lasting stance borrowed from the biological vision research that assumes human-like information processing as an enigmatic mix of perceptual and cognitive vision faculties. I am trying to find a remedy for this bizarre situation. Relying on a new definition of “information”, which can be derived from Kolmogorov’s complexity theory and Chaitin’s notion of algorithmic information, I propose a unifying framework for visual information processing, which explicitly accounts for the perceptual and cognitive image processing peculiarities. I believe that this framework will be useful to overcome the difficulties that are impeding our attempts to develop the right model of human-like intelligent image processing.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Visual information; Physical and semantic image information; Knowledge and information relationship

1. Introduction

The explosive growth of visual information in our surroundings has raised an urgent demand for effective means for organizing and handling these immense volumes of information [1]. Because humans are known to be very efficient in such tasks, it is not surprising that computer vision designers are trying again and again to get answers for their worrying problems among the solutions that human visual system has developed in course of millions of

years of natural evolution. Nearly half of our cerebral cortex is busy with processing visual information [2], but how it is done “in vivo” remains a puzzle for many generations of thinkers, philosophers, and contemporary scientific researchers.

Nevertheless, a working theory of human visual information processing has been established about 25 years ago by the seminal works of Marr [3], Treisman [4], Biederman [5], and a large group of their associates and followers. Since then it has become a classical theory, which dominates today in all further developments in the field. The theory considers human visual information processing as

E-mail address: emanl@012.net.il

an interplay of two inversely directed processing streams. One is an unsupervised, bottom-up directed process of initial image information pieces discovery and localization. The other is a supervised, top-down directed process, which conveys the rules and the knowledge that guide the linking and binding of these disjoint information pieces into perceptually meaningful image objects.

In modern biological vision research, this duality is referred to as perceptual and cognitive faculties of vision. In computer vision terminology, these are the low-level and high-level paths of visual information processing. Although Treisman's theory [4] definitely positions itself as "A Feature-Integration Theory", the difficulties in defining proper rules for this feature integration have impelled a growing divergence between perceptive and cognitive fields of image processing [6]. Obviously, that was a wrong and a counter-productive development, and human vision researchers were always aware of its harmful consequence [7]. For this reason, the so-called "binding problem" has been announced as a critical exploration goal, and massive research efforts have been directed to its resolution [8]. Unfortunately, without any discernable success.

In computer vision, the situation is even more bizarre. Thus far, computer vision community was so busy with its everyday problems that there was no time to raise basic research ventures. Principal ideas as well as their possible solutions are usually borrowed from biological vision research. Therefore, following the trends in biological vision, for decades computer vision R&D has been deeply plunged into bottom-up pixel-oriented image processing. Low-level image computations have become its prime and persistent goal, while the complicated issues of high-level processing were just neglected and disregarded.

However, it is impossible to ignore them completely. It is generally acknowledged that any kind of image processing is unfeasible without incorporation into it the high-level knowledge ingredients. For this reason, the whole history of computer-based image processing is an endless saga on attempts to seize the needed knowledge in any possible way. The oldest and the most common ploy is to capitalize on the domain expert's knowledge and adapt it to each and every application case. It is not surprising, therefore, that the whole realm of image processing has been, and continues to be, fragmented according to the high-level knowledge competence of the experts in the corresponding

domains. That is why we have today: medical imaging, aerospace imaging, infrared, biologic, underwater, geophysics, remote sensing, microscopy, radar, biomedical, X-ray, and so on "imaging".

The advent of the Internet, with huge volumes of visual information scattered over the web, has demolished the long-lasting custom of capitalizing on the expert's knowledge. Image information content on the Web is unpredictable and diversified. It is useless to apply specific expert knowledge to a random set of distant images. To meet the challenge, the computer vision community has undertaken an enterprise to develop the so-called content-based image retrieval (CBIR) technologies [9,10]. However, deprived of any reasonable sources of the desired high-level information, computer vision designers were forced to proceed in only one possible direction of trying to derive the high-level knowledge from the available low-level information pieces [11,12].

In doing so, computer vision designers have once again demonstrated their reliance on biological vision trends and fashions. In biological vision, a rank of theoretical and experimental work has been done in order to support and to justify this above-mentioned tendency. Two ways of thinking could be distinguished in this regard: chaotic attractors modeling [13,14], and saliency attention map modeling [15,16]. We will not review these approaches in details. We will only note that both of them presume low-level bottom-up processing as the most proper way for high-level information recovery. Both are computationally expensive. Both definitely violate the basic assumption about the leading role of high-level knowledge in the low-level information processing.

It will be a mistake to say that computer vision people are not aware of these discrepancies. On the contrary, they are well informed about what is going on in the field. However, they are trying to justify their attempts by promoting a concept of a "semantic gap", an imaginary gap between low- and high-level image features. They sincerely believe that they would be able to bridge it some day [17].

It is worth to mention that all these developments—feature binding in biological vision and semantic gap bridging in computer vision—are evolving in atmosphere of total indifference to prior claims about high-level information superiority in the general course of visual information processing. Such indifference seems to stem from a very loose understanding about what is the concept of

“information”, what is the right way to use it properly, and what information treatment options could arise from this understanding.

2. Re-examining the basic assumptions

Everyone, who is not deaf, knows that we live today in the Information Age, where information is an indispensable ingredient of our life. We consume it, create it, seek it, transfer, exchange, hide, reveal, accumulate, and disseminate it—in one word: information is a remarkably important component of our life. But can someone explain me what we have in mind when the word “information” is uttered? My attempts (undertaken several years ago) to get my own answer for this question were so desperate that I was almost ready to accept the stance that information is an indefinable entity (like “space” and “time” in classical physics, e.g.). Fortunately, at the end, I have hit on an information definition fitting my visual information handling aims. It turns out that this definition can be derived from Solomonoff’s Theory of Inference [18], Chaitin’s Algorithmic Information Theory [19], and Kolmogorov’s Complexity Theory [20]. Recently, I have learned that Kolmogorov’s Complexity and Chaitin’s Algorithmic Information Theory are referred as respected items of a list of seven possible contestants suitable to define what actually information is [21]. In this regard, I was very proud of myself that I was lucky to avoid the traps of Shannon’s Information Theory, which is known to be useful in communication applications, but it is absolutely inappropriate for visual information explorations that I am trying to conduct. The reason for this is that Shannon’s information properly describes the integrated properties of an information message, while Kolmogorov’s definition is suitable for evaluation of information content of separate isolated subparts of a message (separate message objects). This is, indeed, much closer to the way in which humans perceive and grasp their visual information.

The results of my investigation have been already published on several occasions [22–25], and interested readers can easily get them from a number of freely accessible repositories (e.g., arXiv, CiteSeer (the former Research Index), Eprintweb, etc.). Therefore, I will only repeat here some important points of these early publications, which properly reflect my current understanding of the matters.

The main point is that *information is a description*, a certain alphabet-based or language-based description, which Kolmogorov’s theory regards as a program that (when executed) trustworthy reproduces the original object [26]. In an image, such objects are visible data structures from which an image consists of. So, a set of reproducible descriptions of image data structures is the information contained in an image.

The Kolmogorov’s theory prescribes the way in which such descriptions must be created: at first, the most simplified and generalized structure must be described. Then, as the level of generalization is gradually decreased, more and more fine-grained image details (structures) are become revealed and depicted. This is the second important point, which follows from the theory’s pure mathematical considerations: *image information is a hierarchy of recursive decreasing level descriptions* of information details, which unfolds in a coarse-to-fine top-down manner. (Attention please! Any bottom-up processing is not mentioned here. There is no low-level feature gathering and no feature binding. The only proper way for image information elicitation is a top-down coarse-to-fine way of image processing.)

The third prominent point, which immediately pops-up from the two just mentioned above, is that the top-down manner of image *information elicitation does not require incorporation of any high-level knowledge* for its successful accomplishment. It is totally free from any high-level guiding rules and inspirations. What immediately follows from this, is that high-level image semantics is not an integrated part of image information content (as it is traditionally assumed). It cannot be seen more as a natural property of an image. Image semantics must be seen as a property of a human observer that watches and scrutinizes an image. That is why we can say now: *semantics is assigned to an image by a human observer*. That is strongly at variance with the contemporary views on the concept of semantic information. Following the new information elicitation rules, it is impossible to continue to pretend that semantics can be *extracted from an image* (as in [27,28]), or should be *derived from low-level information features* via the semantic gap bridging (as in [29,30], and many others). That simply does not hold any more.

3. Computer vision implications

This new definition of information has forced us to reconsider our former approach to image

information processing. The validity of our new assumptions and the inevitable changes in design philosophy that acceptance of these assumptions imply, have motivated us to test the issues in a framework of a visual robot design enterprise. The enterprise is aimed to creating an artificial vision system with some human-like cognitive capabilities. It is generally agreed that the first stage of such a system has to be an image segmentation stage at which the whole bulk of image pixels (image raw data) has to be decomposed into a finite set of image patches. The latter are submitted afterwards to a process of image content analysis and interpretation.

A practical algorithm based on the announced above principles has been developed and subjected to some systematic evaluations. The results were published, and can be found in [23–25]. There is no need to repeat again and again that excellent, previously unattainable segmentation results have been attained in these tests, undoubtedly corroborating the new information processing principles. Not only an unsupervised segmentation of image content has been achieved (in a top-down coarse-to-fine processing manner, without any involvement of high-level knowledge). A hierarchy of descriptions for each and every segmented lot (segmented object) has been achieved as well. It contains the center of mass coordinates, the direction of object's main axes, object's contour and shape depiction rules (in a system of these axes), and other object related parameters (object related *information*), which enable subsequent object reconstruction. That is exactly what we have previously defined as information. That is the reason why we specify this information as "physical information", because that is the only information present in an image, and therefore *the only information that can be extracted from an image*. For that reason, it must be dissociated from the semantic information, which (as we understand now) is a property of an external observer. Therefore, it must be treated (or modeled) in accordance with specific his/her cognitive information processing rules.

What are these rules? A consensus view on this topic does not exist as yet in the biological vision theories and in the computer vision practice. So, we have to blaze our own trails. We decided, thus, to meet this challenge by suggesting a new approach based on our previously declared information elicitation principles. The preliminary results of our first attempt were published recently in [31]. As

in the case of physical information, we will not repeat here all the details of this publication. We will proceed with only a brief reproduction of some critical points needed to follow up our discussion.

Human's cognitive abilities (including the aptness for image interpretation and the capacity to assign semantics to an image) are empowered by the existence of a huge knowledge base about the things in the surrounding world kept in human brain/head. This knowledge base is permanently upgraded and updated during the human's life span. So, if we intend to endow our visual robot with some cognitive capabilities we have to provide it with something equivalent to this (human) knowledge base.

It goes without saying that this knowledge base will never be as large and developed as its human prototype. But we are not sure that the requirement to be large and huge is valid in our case. After all, humans are also not equal in their cognitive capacity, and the magnitude, the content of their knowledge bases is very diversified too. (The knowledge base of aerial photographs interpreter is certainly different from the knowledge base of roentgen images interpreter, or IVUS images, or PET images.) The knowledge base of our visual robot has to be small enough to be effective and manageable, but sufficiently ample to ensure robot's acceptable performance. Certainly, for our feasibility study we can be satisfied even with a relatively small, specific task-oriented knowledge base.

The next crucial point is the knowledge (base) representation issue. To deal with it, we first of all must arrive at a common agreement about what is the meaning of the term "knowledge". (A question that usually has not a commonly accepted answer.) We state that in our case a suitable and a sufficient definition of it would be: "Knowledge is a memorized information". Consequently, we can say that knowledge (like information) must be a hierarchy of descriptive items, with the grade of description details growing in a top-down manner at the descending levels of the hierarchy.

What else must be mentioned here, is that these descriptions have to be implemented in some alphabet (as it is in the case of physical information) or in a description language (which better fits the semantic information case). Any further argument being put aside, we will declare that the most suitable language in our case is a natural human language. After all, the real knowledge bases that we are familiar with are implemented on a natural human language basis.

The next step, then, is pre-determined: if natural language is a suitable description implement, the suitable form of this implementation is a narrative, a story tale [32]. If the description hierarchy can be seen as an inverted tree, then the branches of this tree are the stories that encapsulate human's experience with the surrounding world. And the leaves of these branches are single words from which the stories are composed of. In computer vision terminology these single words are defined as nodes.

The descent into description details, however, does not stop here, and each single word can be further decomposed into its attributes and rules that describe the relations between the attributes. At this stage the notion of physical information comes back to the game. Because the words are usually associated with physical objects in the real world, words' attributes must be seen as memorized physical information descriptions. Once derived (by a visual system) from the observable world and learned to be associated with a particular word, these physical information descriptions are soldered in into the knowledge base. Object recognition, thus, turns out to be a comparison and similarity test between currently acquired physical information and the one already retained in the memory. If the similarity test is successful, starting from this point in the hierarchy and climbing back on the knowledge base ladder we will obtain: first, the linguistic label for a recognized object, and second, the position of this label (word) in the context of the whole story. In this way, object's meaningful categorization can be acquired, a first stage of image annotation can be successfully accomplished, paving the way for further meaningful (semantic) image interpretation.

One question has remained untouched in our discourse: How this artificial knowledge base has to be initially created and brought into the robot's disposal? The vigilant reader certainly remembers the fierce debates about learning capabilities of neural networks and other machine learning technologies. We are aware of these debates. But in our case they are irrelevant for a simple reason: the top-down fashion of the knowledge base development pre-determines that all responsibilities for knowledge base creation have to be placed on the shoulders of the robot designer.

Such an unexpected twist in design philosophy will be less surprising if we recall that human cognitive memory is also often defined as a

“declarative memory”. And the prime mode of human learning is the declarative learning mode, when the new knowledge is explicitly transferred to a developing human from his external surrounding: From a father to a child, from a teacher to a student, from an instructor to a trainee. There is evidence that this is not an especially human prerogative. Even ants are transferring knowledge in a similar way [33]. So, our proposal that robot's knowledge base has to be designed and created by the robot supervisor is sufficiently correct and is fitting our general concept of information use and management.

4. Brain vision implications

Since the beginning of the computer vision age, there was a common belief that biological vision is an endless source of inspiration for the computer vision designs, and that the main ideas, which underpin contemporary computer vision implementations, are borrowed from the fruitful fields of biological vision research. In Section 1 of this paper, we have shown that such farfetched credits are groundless. In the course of time, only the bottom-up processing philosophy has remained as a common feature that still can be found both in biological and in computer vision designs. However, practical implementations of this philosophy have led to very different and incompatible developments: massive DSP-based parallel processing in computer vision and selective attention-based sequential processing in biological vision. This divergence is an annoying misfortune. In [25], I tried to allege that there must be a general underpinning basis able to reconcile the this day detached subdivisions. I believe that the unified information processing framework proposed in this paper can be useful in pursuing this goal. In this regard, it is tempting to see how biological vision research can benefit from the proposed new ideas and to what extent elements of these new ideas can be discerned in the ongoing biological vision experiments.

While the mainstream of biological vision research continues to approach visual information processing in a bottom-up fashion [34], it turns out that the idea of primary top-down processing was never been extraneous to biological vision. The first publications addressing this issue are dated by the early eighties of the last century (David Navon at 1977 [35], and Lin Chen at 1982 [36]). The prominent authors were persistent in their views

on the matters, and further research reports have been published regularly until the recent time [37,38]. However, it looks like they all have been overlooked, both in the biological and in the computer vision. Only in the last years, a tide of new evidence has appeared that definitely advocates for the top-down processing primordially [39,40].

The field of cognitive vision is not ready yet to leave the traditional information processing dogmas. However, supporting evidence for a “declarative” interpretation of physical information can be already found in [41], where it is convincingly shown how a color is being “assigned” to an object.

Knowledge transfer from the outside and information description exchange in course of knowledge base building and declarative learning accomplishment have been observed not only in humans. They are ubiquitous even among far more primitive living beings: ants that learn in tandem [33] we have already mentioned in Section 3. The so-called horizontal gene transfer responsible for antibiotic resistance development of bacteria [42] can also be seen as a supporting evidence for a top-down (external) knowledge base creation phenomenon.

However, the most surprising insights are still awaiting their further clarification and confirmation. If our definition of information as a description is correct, then the current belief that a spiking neuron burst is a valid form of information representation and exchange [43] does not hold any more. Variances in spikes’ heights or duty times are not an adequate alphabet to implement information descriptions of a desired complexity. We can boldly speculate that a biomolecular alphabet would be a much better and appropriate solution for such cases. Support for this kind of speculations can be derived from the recent advances in molecular biology research [44,45]. The spikes that we observe and investigate today could be seen as a reflection of charges that are carried by the ionized parts of the molecular information messages.

This molecular description hypothesis fits very well also our new brain memory organization theory, which pretty well resembles the paradigm of computer memory use. Dendrite spines can be seen as a proper accommodation for the molecular descriptors, providing a kind of a biological hardware, a biological memory buffer, which is able to hold read/write-able information messages. Computer inspired hypotheses about “object files” [46] and “event files” [47] are repeatedly emerging in biological vision literature over the last decades. It

would be interesting to notice that the concept of a biological event file fits very well also the narrative knowledge representation hypothesis proposed earlier in Section 3.

Further support for the idea that a complex information description can be stored in a single neuron memory cell can be seen in [48]. In this case as well, reactivation and retrieval of a memorized description resembles well the paradigm of a computer memory bank store/fetch transaction, a single_write/multiple_read memory access operation.

I am definitely excited by the options that brain vision research can gain from such a back projection of a computer vision theory (about the essence of information) on the issues of modeling visual information processing in human brain.

5. Some concluding remarks

In this paper, we propose a new definition of information, in particular, a definition of visual information, which is the prime point of our concern. Relying on this definition, a substantial progress has been achieved in modeling human-like intelligent image processing.

The approach outlined in this paper appears to conflict with the mainstream of research and development enterprises undertaken in the frame of European IST and USA TRECVID initiatives [49,50]. The main points of divergence are as follows. The bottom-up way of image processing (from local details to more general forms of representation), commonly accepted today as the preferred processing alternative, is applicable only if the similarity metrics can be somehow defined a priori. Generally, that is not the case. The right way of visual information processing is a top-down (from coarse-to-fine) processing alternative, as it is proposed in this paper.

Relying on the bottom-up processing, the right structure of the information hierarchy can never be caught properly. That is the source of many deadfalls that the mainstream affiliates experience over and over again. For example, the unlucky attempts to solve the semantic gap problem stem from an improper placement of perceptual and cognitive vision features on the same hierarchical level. The same bottom-up preferences perpetually derail the machine learning-based approaches for knowledge/information acquisition and handling. From a top-down processing perspective, a

“machine teaching” approach, when knowledge is brought into from the outside, looks as a far more appropriate solution.

The examples mentioned above do not exhaust the list of changes that must be imposed on the mainstream projects if the new information processing rules are taken into the consideration. It is obvious that these rules are incomplete and tentative, since this is just a first step, and further research remains to be done. The enterprise that we are aimed at, is not a task for a single person or a small group of developers. It requires consolidated efforts of many interesting parties. We hope that the time for this collaboration is not far away.

References

- [1] M.L. Khefri, D. Ziou, A. Bernardi, Image retrieval from the world wide web: issues, techniques, and systems, *ACM Comput. Surveys* 36 (1) (March 2004) 35–67.
- [2] D. Milner, M. Goodale, *The Visual Brain in Action*, Oxford Psychology Series, 27, Oxford University Press, Oxford, 1998.
- [3] D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, Freeman, San Francisco, 1982.
- [4] A. Treisman, G. Gelade, A feature-integration theory of attention, *Cogn. Psychol.* 12 (January 1980) 97–136.
- [5] I. Biederman, Recognition-by-components: a theory of human image understanding, *Psychol. Rev.* 94 (2) (1987) 115–147.
- [6] L.W. Barsalou, Perceptual symbol systems, *Behav. Brain Sci.* 22 (1999) 577–660.
- [7] T. Palmeri, I. Gauthier, Visual object understanding, *Nat. Rev.: Neurosci.* 5 (April 2004) 291–304.
- [8] A. Treisman, The binding problem, *Curr. Opin. Neurobiol.* 6 (1996) 171–178.
- [9] S.-F. Chang, W.-Y. Ma, A. Smeulders, Recent advances and challenges of semantic image/video search, in: *Proceedings of the IEEE Int. Conference on Acoustics, Speech, and Signal Processing, ICASSP-2007*. Available on: <http://staff.science.uva.nl/~smeulder/>.
- [10] M.S. Lew, N. Sebe, C. Djeraba, R. Jain, Content-based multimedia information retrieval: state of the art and challenges, *ACM Trans. Multimedia Comput. Commun. Appl.* (February 2006).
- [11] A. Mojsilovic, B. Rogowitz, Capturing image semantics with low-level descriptors, in: *Proceedings of the International Conference on Image Processing (ICIP-01)*, Thessaloniki, Greece, October 2001, pp. 18–21.
- [12] C. Zhang, T. Chen, From low level features to high level semantics, in: *Handbook of Video Databases: Design and Applications*, by Furht, Borko/Marques, Oge, CRC Press, October 2003.
- [13] K. McRae, Semantic memory: some insights from feature-based connectionist attractor networks, in: B.H. Ross (Ed.), *The Psychology of Learning and Motivation*, vol. 45, 2004, available at: <http://amdrae.ssc.uwo.ca/>.
- [14] C. Johansson, A. Lansner, Attractor memory with self-organizing input, in: *Workshop on Biologically Inspired Approaches to Advanced Information Technology (BioA-DIT 2005)*, *Lecturer Notes in Computer Science*, vol. 3853, Springer, 2006, pp. 265–280.
- [15] S. Treue, Visual attention: the where, what, how and why of saliency, *Curr. Opin. Neurobiol.* 13 (2003) 428–432.
- [16] L. Itti, G. Rees, J. Tsotsos (Eds.), *Neurobiology of Attention*, San Diego, CA, Elsevier, 2005, pp. 576–582.
- [17] J. Hare, P. Lewis, P. Enser, C. Sandom, Mind the gap: another look at the problem of the semantic gap in image retrieval, in: *Proceedings of Multimedia Content Analysis, Management and Retrieval Conference, SPIE* vol. 6073, 2006, available at: <http://www.ecs.soton.ac.uk/people/>.
- [18] R.J. Solomonoff, The discovery of algorithmic probability, *J. Comput. Syst. Sci.* 55 (1) (1997) 73–88.
- [19] G.J. Chaitin, Algorithmic information theory, *IBM J. Res. Dev.* 21 (1977) 350–359.
- [20] A. Kolmogorov, Three approaches to the quantitative definition of information, *Probl. Inf. Transm.* 1 (1) (1965) 1–7.
- [21] L. Floridi, Trends in the philosophy of information, in: P. Adriaans, J. van Benthem (Eds.), *Handbook of Philosophy of Information*, Elsevier, forthcoming, Available at: <http://www.philosophyofinformation.net>.
- [22] E. Diamant, Image information content estimation and elicitation, *WSEAS Trans. Comput.* 2 (2) (2003) 443–448 <http://www.worldses.org/journals/>.
- [23] E. Diamant, Top-down unsupervised image segmentation (it sounds like an oxymoron, but actually it isn't), in: *Proceedings of the third Pattern Recognition in Remote Sensing Workshop (PRRS'04)*, August 2004, Kingston University, UK.
- [24] E. Diamant, Searching for image information content, its discovery, extraction, and representation, *J. Electron. Imaging* 14 (1) (2005).
- [25] E. Diamant, Does a plane imitate a bird? Does computer vision have to follow biological paradigms? in: M. De Gregorio, et al. (Eds.), *Brain, Vision, and Artificial Intelligence, First International Symposium Proceedings*, *Lecturer Notes in Computer Science*, vol. 3704, Springer, pp. 108–115, 2005, Available at: <http://www.vidiament.info>.
- [26] P. Vitanyi, Meaningful information, *IEEE Trans. Inf. Theory* 52 (10) (October 2006) 4617–4624 Available at: <http://www.cwi.nl/~paulv/papers>.
- [27] M. Naphade, T.S. Huang, Extracting semantics from audiovisual content: the final frontier in multimedia retrieval, *IEEE Trans. Neural Networks* 13 (4) (July 2002) 793–810.
- [28] J. Jiten, B. Meriardo, B. Huet, Semantic feature extraction with multidimensional hidden Markov model, in: *SPIE Conference on Multimedia Content Analysis, Management and Retrieval*, 17–19 January 2006, San Jose, USA, SPIE vol. 6073, pp. 211–221, 2006.
- [29] X.S. Zhou, T.S. Huang, CBIR: from low-level features to high-level semantics, in: *Proceedings of the SPIE*, vol. 3974, San Jose, CA, 24–28 January 2000, pp. 426–431, available at: <http://www.ifp.uiuc.edu/~xzhou/>.
- [30] K. Petridis, I. Kompatsiaris, M. Strintzis, S. Bloehdorn, S. Handschuh, S. Staab, N. Simou, V. Tzouvaras, Y. Avrithis, Knowledge representation for semantic multimedia content

- analysis and reasoning, *IEEE Trans. CSVT* 15 (10) (October 2005) 1210–1244 Available at: <http://www.iti.gr/db.php/publications>.
- [31] E. Diamant, In Quest of Image Semantics: Are We Looking for it under the Right Lamppost? available at: <http://arxiv.org/abs/cs.CV/0609003>.
- [32] M. Tuffield, N. Shadbolt, D. Millard, Narratives as a form of knowledge transfer: narrative theory and semantics, in: Proceedings of the 1st AKT (Advance Knowledge Technologies) Symposium, Milton Keynes, UK, June 2005.
- [33] N. Franks, T. Richardson, Teaching in tandem-running ants, *Nature* 439 (12 January 2006) 153.
- [34] T. Serre, M. Kouh, C. Cadieu, U. Knoblich, G. Kreiman, T. Poggio, A Theory of Object Recognition: Computations and Circuits in the Feedforward Path of the Ventral Stream in Primate Visual Cortex, AI Memo 2005-036/CBCL Memo 259, MIT, Cambridge, November 2005, available at: <http://web.mit.edu/serre/publications/...>.
- [35] D. Navon, Forest before trees: the precedence of global features in visual perception, *Cogn. Psychol.* 9 (1977) 353–383.
- [36] L. Chen, Topological structure in visual perception, *Science* 218 (1982) 699–700.
- [37] D. Navon, What does a compound letter tell the psychologist's mind?, *Acta Psychol.* 114 (2003) 273–309.
- [38] L. Chen, The topological approach to perceptual organization, *Vis. Cogn.* 12 (4) (2005) 553–637.
- [39] M. Ahissar, S. Hochstein, The reverse hierarchy theory of visual perceptual learning, *Trends Cogn. Sci.* 8 (10) (2004) 457–464.
- [40] C.-H. Juan, G. Campana, V. Walsh, Cortical interactions in vision and awareness: hierarchies in reverse, *Prog. Brain Res.* 144 (2004) 117–130.
- [41] T. Hansen, M. Olkkonen, S. Walter, K. Gegenfurtner, Memory modulates color appearance, *Nat. Neurosci.* 9 (11) (November 2006) 1367–1368.
- [42] J. Lawrence, H. Hendrickson, Lateral gene transfer: when will adolescence end?, *Mol. Microbiol.* 50 (3) (2003) 739–749.
- [43] K. Diba, C. Koch, I. Segev, Spike propagation in dendrites with stochastic ion channels, *J. Comput. Neurosci.* 20 (2006) 77–84.
- [44] E. Kandel, The molecular biology of memory storage: a dialogue between genes and synapses, *Science* 294 (2 November 2001) 1030–1038.
- [45] A. Routtenberg, J. Rekart, Post-translational protein modification as the substrate for long-lasting memory, *Trends Neurosci.* 28 (1) (January 2005) 12–19.
- [46] D. Kahneman, A. Treisman, B. Gibbs, The reviewing of object files: object-specific integration of information, *Cogn. Psychol.* 24 (1992) 175–219.
- [47] B. Hommel, Event files: feature binding in and across perception and action, *Trends Cogn. Sci.* 8 (11) (2004) 494–500.
- [48] S. Waydo, A. Kraskov, R. Quiroga, I. Freid, C. Koch, Sparse representation in the human medial temporal lobe, *J. Neurosci.* 26 (40) (4 October 2006) 10232–10234.
- [49] European IST Research (2005–2006): Building on Assets, Seizing Opportunities, available at: http://europa.eu.int/information_society/.
- [50] P. Over, T. Ianeva, W. Kraaij, A. Smeaton, TRECVID 2005—An Overview, available at: <http://www.uv.es/~tzveta/>.